A stratified log-rank test with missing stratum information and dependent censoring

Amel Mezaouer, Jean-François Dupuy, Kamal Boukhetala

A stratified log-rank test with missing stratum information and dependent censoring

A. Mezaouer

Faculté des Sciences, Université Saad Dahleb, Blida, Algeria.
Email: a_mezaouer@univ-blida.dz

J.-F. Dupuy†

IRMAR-Institut National des Sciences Appliquées de Rennes, France.
Email: Jean-Francois.Dupuy@insa-rennes.fr

K. Boukhetala

Université des Sciences et Technologie Houari Boumédienne, Bab-Ezzouar, Algeria.
Email: kboukhetala@usthb.dz

Abstract. The log-rank test is often used to compare randomized treatment groups with respect to the distribution of a failure time outcome. The so-called stratified log-rank test can be used when it is necessary to adjust for the effect of some discrete covariate that may be predictive of the outcome. In many applied situations, this discrete covariate is missing for some of the patients and moreover, the distribution of the censoring time depends on the treatment group. In this paper, we introduce a modified version of the stratified log-rank test, which accommodates both these problems simultaneously. The asymptotic distribution of this new test under the null hypothesis of equality of the randomized treatment groups is established. A numerical study is conducted to examine the finite-sample behavior of this test under both the null and alternative hypotheses.

1. Introduction

The log-rank test is widely used to compare randomized treatment groups with respect to the distribution of some failure time outcome. If one needs to control for a covariate that may be predictive of the outcome, and if this covariate is discrete, the log-rank can be generalized to the so-called stratified log-rank test (see, for example, Klein and Moeschberger (1997) and Martinussen and Scheike (2006)). Consider a clinical trial where n patients are randomly assigned to K different treatment groups. We wish to compare survival between groups, while adjusting for some discrete factor S with L modalities (also called strata, such as income groups or disease stages for example). If $\lambda_{k,l}$ is the instantaneous hazard function for a patient in the kth treatment group and lth stratum, then the test for treatment effect can be formulated as

$$H_0 : \lambda_{1,l} = \ldots = \lambda_{K,l} \quad \text{for every } l = 1, \ldots, L$$

versus H_a: "there exists j and j' such that $\lambda_{j,l} \neq \lambda_{j',l}$ for some l". Let $T_1^0, \ldots, T_n^0$ be the times from randomization to failure observed in the K pooled groups. Let $C_1, \ldots, C_n$ be right-censoring times (the C_i are assumed to be independent of the T_i^0 and non-informative).

†Corresponding author

For each patient i, we observe $T_i = \min(T_i^0, C_i)$ and $\Delta_i = 1(T_i^0 \leq C_i)$, where $1(\cdot)$ is the indicator function. Assume that the data consist of n independent and identically distributed quadruplets $(T_i, \Delta_i, G_i, S_i)$, $i = 1, \ldots, n$, where $G_i \in \{1, \ldots, K\}$ and $S_i \in \{1, \ldots, L\}$ respectively indicate the group and stratum of the ith patient. Let $N_i(t) = \Delta_i 1(T_i \leq t)$, $Y_i(t) = 1(T_i \geq t)$, and define

$$E_{k,l}^{(n)}(t) = \frac{\sum_{i=1}^n Y_i(t) 1(G_i = k) 1(S_i = l)}{\sum_{i=1}^n Y_i(t) 1(S_i = l)}.$$

Then the stratified log-rank statistic for the test of no randomized treatment effect is of the form

$$U = (Z_1, \ldots, Z_{K-1}) \hat{\Theta}^{-1} (Z_1, \ldots, Z_{K-1})',$$

where for every $k = 1, \ldots, K-1$,

$$Z_k = \sum_{i=1}^n \int_0^\tau \left\{ 1(G_i = k) - \sum_{l=1}^L 1(S_i = l) E_{k,l}^{(n)}(t) \right\} dN_i(t), \tag{1}$$

τ denotes the end of the study period, and $\hat{\Theta}$ is the estimated asymptotic covariance matrix of $(Z_1, \ldots, Z_{K-1})'$. Under H_0, U is asymptotically distributed as a χ^2 distribution with $K-1$ degrees of freedom (see, for example, Martinussen and Scheike (2006)).

In some applications, the stratum S may be missing for some patients. For example, consider the case where S represents the histological stage of patients included in a cancer clinical trial. The determination of S may require a biopsy, which due to expensiveness may not be performed on all the study subjects. One simple solution to handle such incomplete stratum information is to perform a complete-case analysis that is, to discard patients with unobserved stratum. This, however, may induce a substantial loss of power, as will be illustrated in our simulation study. Dupuy and Leconte (2008, 2009) considered the distinct but related problem of estimation in the stratified proportional hazards model with missing strata. The authors proposed a modified version of the maximum partial likelihood estimator, in which the unobserved stratum indicators are replaced by an estimate of their conditional expectation given available auxiliary covariates (this is the so-called regression calibration idea, see for example Carroll *et al.* (1995)). Simulation results provided some evidence that such a replacement substantially improves on the complete-case analysis.

In many applications, it also happens that the distribution of censoring time depends on the treatment group. This arises, for example, when censoring follows from a study dropout caused by treatment toxicity. The treatment group with the heaviest toxicity will be more likely to have a higher dropout rate, and thus a higher censoring rate, than the other groups. Inverse probability of censoring weighted (IPCW) procedures have been proposed to remedy this problem (see for example Robins and Finkelstein (2000), Yoshida *et al.* (2007), Cain and Cole (2009)).

In this paper, we propose and investigate a test of no randomized treatment effect, when the patients stratum information is only partially available *and* the distribution of censoring time depends on the treatment group. The test we propose combines the regression calibration and IPCW principles.

The rest of the article is organized as follows. In Section 2, we introduce the new test statistic and we derive its asymptotic distribution under H_0. In Section 3, we describe a short simulation study investigating the finite-sample behaviour of the proposed test. A discussion concludes the paper in Section 4. An appendix contains the proofs of some intermediate technical results.

2. The test statistic

Assume that n independent patients are randomly assigned to K treatment groups. Assume that the stratum value is missing for some of these patients. Thus, a subsample is available where all variables (T, Δ, G, S) are observed, while only (T, Δ, G) are observed for the other patients. We assume that some auxiliary variables $W \in \mathbb{R}^p$ are observed for all patients, and that W provides a partial information about S when S is missing. Let R be the indicator variable which is 1 if S is observed and 0 otherwise. Throughout the paper, we assume that T^0 and C are independent given G, S, W and R, and that C is independent of S and W given G. However as mentioned above, the distribution of the censoring time depends on the treatment group. We assume that T^0 is independent of W given S (that is, the auxiliary variables W provide no additional information about failure when the true stratum S is known), and that G is independent of S and W, as is the case in randomized clinical trials. We assume that R is independent of T^0, C and G, and that $0 < \mathbb{P}(R = 1) < 1$. Finally, we assume that R and S are independent given W, which is the so-called missing-at-random assumption. In the sequel, this set of assumptions will be denoted by C1.

We consider the problem of implementing the stratified log-rank test of H_0 based on n independent vectors $(T_i, \Delta_i, G_i, W_i, R_i, R_iS_i)$, $i = 1, \ldots, n$ of possibly incomplete data when moreover, the distribution of censoring time depends on the treatment group. To tackle simultaneously the missing strata and dependent censoring problems, we introduce a modified version of U, which is obtained by:

(a) replacing any missing stratum indicator $1(S_i = l)$ in (1) by its conditional expectation given the auxiliary W (this idea is related to regression calibration methods, see, for example, Thurston *et al.* (2003), Weller *et al.* (2007), Dupuy and Leconte (2009)), and

(b) weighting every patient by the inverse of the conditional (given the patient's treatment group) survival function of the censoring time (this idea is related to the inverse probability of censoring weighted principle, e.g., Robins and Finkelstein (2000), Yoshida *et al.* (2007), Cain and Cole (2009)).

Precisely, we propose to base our test statistic on the following modified version of (1):

$$\widetilde{Z}_k = \sum_{i=1}^{n} \int_0^{\tau} \mu(G_i, t) \left\{ G_i^k - \sum_{l=1}^{L} D_i^l \widetilde{E}_{k,l}^{(n)}(t) \right\} dN_i(t), \tag{2}$$

where for every $i = 1, \ldots, n$, $k = 1, \ldots, K$, $l = 1, \ldots, L$, and $t \in [0, \tau]$, $G_i^k = 1(G_i = k)$, $D_i^l = R_i 1(S_i = l) + (1 - R_i)\mathbb{E}[1(S_i = l)|W_i]$, $\mu(G_i, t) = 1/\mathbb{P}(C \geq t|G_i)$ (where $\mathbb{P}(C \geq t|G)$ is the survival function of the censoring time in group G), and $\widetilde{E}_{k,l}^{(n)}(t) = \widetilde{S}_{k,l}^{(n)}(t)/\widetilde{S}_l^{(n)}(t)$, with

$$\widetilde{S}_{k,l}^{(n)}(t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) G_i^k D_i^l \mu(G_i, t) \text{ and } \widetilde{S}_l^{(n)}(t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) D_i^l \mu(G_i, t).$$

Before stating our result, we need to introduce some further notations and some regularity conditions. For every $i = 1, \ldots, n$ and $k = 1, \ldots, K$, let

$$V_{i,k} = \int_0^\tau \mu(G_i,t)\left\{G_i^k - \sum_{l=1}^L D_i^l \widetilde{E}_{k,l}^{(n)}(t)\right\} dN_i(t)$$
$$-\frac{1}{n}\sum_{j=1}^n \sum_{l=1}^L \int_0^\tau \frac{Y_i(t) D_i^l D_j^l \mu(G_i,t)\mu(G_j,t)}{\widetilde{S}_l^{(n)}(t)}\left\{G_i^k - \widetilde{E}_{k,l}^{(n)}(t)\right\} dN_j(t) \qquad (3)$$

and define $\mathbb{V}_i = (V_{i,1}, \ldots, V_{i,K-1})'$ and $\widehat{\Sigma} = \sum_{i=1}^n \mathbb{V}_i \mathbb{V}_i'$. For every $k = 1, \ldots, K$, $l = 1, \ldots, L$, and $t \in [0,\tau]$, define

$$\widetilde{s}_{k,l}(t) = \mathbb{E}[Y(t) G^k D^l \mu(G,t)], \quad \widetilde{s}_l(t) = \mathbb{E}[Y(t) D^l \mu(G,t)],$$

and let $\widetilde{e}_{k,l}(t) = \widetilde{s}_{k,l}(t)/\widetilde{s}_l(t)$. The following regularity conditions will be needed in the proofs:

C2 There exists a positive constant c_0 such that $\mathbb{P}(C \geq \tau | G) > c_0$ for every $G \in \{1, \ldots, K\}$, and the survival function $t \longmapsto \mathbb{P}(C \geq t|G)$ is continuous on $[0,\tau]$.

C3 For every $k = 1, \ldots, K$ and $l = 1, \ldots, L$, $\sup_{t\in[0,\tau]} \lambda_{k,l}(t) < c_1$ for some finite positive constant c_1.

C4 There exists a positive constant c_2 such that $\inf_{t\in[0,\tau]} \widetilde{s}_l(t) > c_2$ for every $l = 1, \ldots, L$.

We are now in position to state the following result for the new test statistic

$$\widetilde{U} := (\widetilde{Z}_1, \ldots, \widetilde{Z}_{K-1})\widehat{\Sigma}^{-1}(\widetilde{Z}_1, \ldots, \widetilde{Z}_{K-1})'.$$

THEOREM 1. *Assume conditions C1-C4. Then under H_0, as $n \to \infty$, $\widetilde{U}$ converges in distribution to a χ^2 distribution with $K-1$ degrees of freedom.*

Based on this theorem, the proposed test rejects H_0 if $\widetilde{U} \geq \chi^2_{1-\alpha}(K-1)$, where $\chi^2_{1-\alpha}(K-1)$ is the quantile of order $1-\alpha$ of $\chi^2(K-1)$.

Remark. The key to derive the null asymptotic distribution of the usual stratified log-rank test U is to represent Z_k as a martingale process and to use a central limit theorem for martingales (see Fleming and Harrington (1991) for example). This is not possible in our case, since the sum across strata in Z_k has been replaced, in $\widetilde{Z}_k$, by a sum where each patient i such that $R_i = 0$ contributes to each of the strata. Therefore, in order to obtain the null asymptotic distribution of $\widetilde{U}$, we rather prove that $\widetilde{Z}_k$ is asymptotically linear, and we use a central limit theorem for sums of i.i.d. terms.

Proof of Theorem 1.

Assume that all random variables are defined on a probability space $(\Omega, \mathcal{C}, \mathbb{P})$, and define $\mathcal{F}_{t,i} = \sigma\{N_i(s), (1-\Delta_i)1(T_i \leq s), G_i, S_i, W_i : 0 \leq s \leq t\}$ as the σ-algebra generated by the event time and censoring histories of the ith patient over $[0,t]$, and by the group, stratum, and auxiliary informations for this patient. Then by assumption, the $\mathcal{F}_{t,i}$-intensity of the counting process $N_i(t)$ is given by

$$Y_i(t)\lambda_i(t) = Y_i(t)\sum_{k=1}^K \sum_{l=1}^L \lambda_{k,l}(t) G_i^k 1(S_i = l).$$

If S_i is missing, the information for the ith patient is represented by the smaller σ-algebra $\mathcal{G}_{t,i} = \sigma\{N_i(s), (1-\Delta_i)1(T_i \le s), G_i, W_i : 0 \le s \le t\} \subseteq \mathcal{F}_{t,i}$. By the innovation theorem, the intensity of $N_i(t)$ with respect to $\mathcal{G}_{t,i}$ is $Y_i(t)\gamma_i(t) := \mathbb{E}[Y_i(t)\lambda_i(t)|\mathcal{G}_{t-,i}]$, where $\gamma_i(t) = \sum_{k=1}^K \sum_{l=1}^L \lambda_{k,l}(t) G_i^k \mathbb{E}[1(S_i = l)|\mathcal{G}_{t-,i}]$. Finally, letting $\mathcal{H}_{t,i} = (\mathcal{F}_{t,i})^{R_i}(\mathcal{G}_{t,i})^{1-R_i}$ be the observed filtration, $N_i(t)$ has intensity $Y_i(t)\zeta_i(t) := Y_i(t)[\lambda_i(t)R_i + \gamma_i(t)(1-R_i)]$ with respect to $\mathcal{H}_{t,i}$. It follows that $M_i(t) = N_i(t) - \int_0^t Y_i(s)\zeta_i(s)\,ds$ is a martingale with respect to $(\mathcal{H}_{t,i})_{t\ge 0}$. In the sequel, we shall note $\kappa_l(t) = \mathbb{E}[Y(t)\zeta(t)D^l\mu(G,t)]$, $l = 1,\ldots,L$ (note that under the conditions stated above, $\kappa_l(t) < \infty$ for every t).

The following lemma establishes a useful approximation of $n^{-\frac{1}{2}}\widetilde{Z}_k$. Its proof is given in the Appendix A.

Lemma 1. *For every $i = 1,\ldots,n$ and $k = 1,\ldots,K$, let*

$$Q_{i,k} = \int_0^\tau \sum_{l=1}^L D_i^l \mu(G_i,t)\left(G_i^k - \widetilde{e}_{k,l}(t)\right)\left[dN_i(t) - Y_i(t)\frac{\kappa_l(t)}{\widetilde{s}_l(t)}\,dt\right].$$

Then under C1-C4, $n^{-\frac{1}{2}}\widetilde{Z}_k = n^{-\frac{1}{2}}\sum_{i=1}^n Q_{i,k} + o_p(1)$. Moreover, if H_0 holds, $\mathbb{E}[Q_{i,k}] = 0$.

It follows from Lemma 1, from the multivariate central limit theorem, and Slutsky's theorem that under H_0, $n^{-\frac{1}{2}}(\widetilde{Z}_1,\ldots,\widetilde{Z}_{K-1})'$ converges in distribution, as $n\to\infty$, to a $(K-1)$-dimensional Gaussian vector with mean 0 and covariance matrix $\Sigma = \mathbb{E}[\mathbb{Q}_1\mathbb{Q}_1']$, where $\mathbb{Q}_i = (Q_{i,1},\ldots,Q_{i,K-1})'$. Consequently, under H_0, $n^{-1}(\widetilde{Z}_1,\ldots,\widetilde{Z}_{K-1})\Sigma^{-1}(\widetilde{Z}_1,\ldots,\widetilde{Z}_{K-1})' \xrightarrow{d} \chi^2(K-1)$ as $n\to\infty$. Σ however involves several unknown expectations. A consistent estimator of Σ is $n^{-1}\widehat{\Sigma} := n^{-1}\sum_{i=1}^n \mathbb{V}_i\mathbb{V}_i'$, where $\mathbb{V}_i = (V_{i,1},\ldots,V_{i,K-1})'$ and $V_{i,k}$ is given by (3). To see this, since $n^{-1}\widehat{\Sigma} = n^{-1}\sum_{i=1}^n(\mathbb{V}_i\mathbb{V}_i' - \mathbb{Q}_i\mathbb{Q}_i') + n^{-1}\sum_{i=1}^n \mathbb{Q}_i\mathbb{Q}_i'$, it is sufficient to prove that $\mathbb{V}_{i,k} - \mathbb{Q}_{i,k} \xrightarrow{p} 0$ as $n\to\infty$. This follows from similar arguments and calculations as in the detailed proof of Lemma 2 (Appendix B). The details are therefore omitted. Finally, it follows from Slutsky's theorem that $\widetilde{U} := (\widetilde{Z}_1,\ldots,\widetilde{Z}_{K-1})\widehat{\Sigma}^{-1}(\widetilde{Z}_1,\ldots,\widetilde{Z}_{K-1})' \xrightarrow{d} \chi^2(K-1)$ as $n\to\infty$.

□

Remark. In practice, the weighting functions $\mu(G_i,\cdot)$ and/or the conditional probabilities $\mathbb{E}[1(S_i = l)|W_i]$ may be either known from previous studies or completely unknown. In this latter case, estimated functions and probabilities have to be substituted in $\widetilde{U}$. We come back to this issue in the simulation study.

3. A simulation study

We conducted a simulation study to evaluate the small to large-sample size behavior of the proposed test under various conditions. We considered the case of $K = 2$ randomized treatment groups and $L = 2$ strata. In each group and stratum, the event times T_i^0 were generated from a Weibull distribution $W(\alpha,\lambda)$ with hazard rate $\lambda(t) = \alpha\lambda t^{\alpha-1}$ (the Weibull distribution is flexible and has a wide range of applications in survival analysis, see Klein and Moeschberger (1997) for example). The failure times of stratum 1 in group 1 were generated from $W(\alpha_1,\lambda_1)$, and those of stratum 1 in group 2 from $W(\alpha_1,\lambda_1 r_1)$, where '$r_1$' denotes the hazard rates ratio of two patients in stratum 1 of groups 1 and 2 respectively. The failure times of stratum 2 in group 1 were generated from $W(\alpha_2, \lambda_2)$ and those of stratum 2 in group 2 from $W(\alpha_2, \lambda_2 r_2)$, with '$r_2$' being the hazard rates ratio of two patients in stratum 2 of groups 1 and 2 respectively.

We used $\alpha_1 = .5, \alpha_2 = .75, \lambda_1 = .75$, and $\lambda_2 = 1.5$. Three cases were considered for the pairs (r_1, r_2) of hazard ratios: (a) $(r_1, r_2) = (1, 1)$, (b) $(r_1, r_2) = (1.5, 1.5)$, (c) $(r_1, r_2) = (1.25, 2)$. Case (a) corresponds to the null case of no difference between treatment groups, within each stratum. Cases (b) and (c) correspond to various magnitudes of difference between groups. In each case, the censoring times were generated from exponential distributions with parameters θ_1 in group 1 and θ_2 in group 2, with θ_1 and θ_2 chosen to yield censoring percentages equal to c_1 in group 1 and c_2 in group 2 (letting $\theta_1 \neq \theta_2$ ensures that the distribution of censoring depends on the treatment group).

Let n_1 and n_2 denote respectively the sample size in group 1 and 2 (with $n = n_1 + n_2$). We considered various values for (n_1, n_2): $(n_1, n_2) = (50, 50)$, $(n_1, n_2) = (100, 100)$, and $(n_1, n_2) = (150, 150)$. The auxiliary variable W was taken to be 2-dimensional ($W = (W_1, W_2)'$) with W_1 (respectively W_2) generated from the uniform distribution on $[-1, 1]$ (respectively the normal distribution with mean 0 and standard deviation 0.5). A logistic regression model

$$\mathbb{P}(S = 1|W) = \frac{\exp(b_0 + b_1 W_1 + b_2 W_2^2)}{1 + \exp(b_0 + b_1 W_1 + b_2 W_2^2)}$$

was taken for the relationship between S and W, with (b_0, b_1, b_2) chosen so that within each treatment group, each stratum contains approximately half of the patients.

The following stratum missingness percentages were considered: $20\%, 40\%$. For each patient, the missingness indicator R was obtained by randomly drawing a Bernoulli random variable, with parameter chosen to yield the prescribed overall missingness percentage. The design parameters and their values are summarized in Table 1.

Table 1. Design parameters and values included in the simulations

Parameter	Values	Description
(r_1, r_2)	(1,1), (1.5,1.5), (1.25,2)	Ratios of hazard rates
(c_1, c_2)	(5,20), (20,50), (30,20), (40,10)	Censoring percentages
n	100, 200, 300	Total sample size

As mentioned previously, in practical situations the weighting functions $\mu(G_i, \cdot)$ and/or the conditional probabilities $\mathbb{P}(S_i = l|W_i) = \mathbb{E}[1(S_i = l)|W_i]$ may either be known (from previous studies, for example), or completely unknown. In this latter case, one would estimate them and substitute the estimated values in the proposed test statistic $\widetilde{U}$ (the resulting statistic will be denoted by $\widehat{U}$ in the sequel). However, the null asymptotic distribution of $\widehat{U}$ may be somewhat distorted from the $\chi^2(K-1)$ distribution. This, in turn, may affect the size and power of the test based on $\widehat{U}$. In fact, our simulation results show that as long as the $\mu(G_i, \cdot)$ and $\mathbb{P}(S_i = l|W_i)$ are reasonably estimated, the prescribed level of the test is nearly maintained by $\widehat{U}$, and that $\widehat{U}$ outperforms the complete-case log-rank test in term of power.

Various methods may be used to estimate $\mathbb{P}(S_i = l|W_i)$. First, one may assume that there exists a known function h depending on an unknown parameter $\mathbf{b}$, such that $\mathbb{P}(S_i = l|W_i) = h(W_i, \mathbf{b})$. Under the missing-at-random assumption, $\mathbb{P}(S = l|W, R = 0) = \mathbb{P}(S = l|W, R = 1) = \mathbb{P}(S = l|W)$. One may thus obtain a consistent estimator $\widehat{\mathbf{b}}$ of $\mathbf{b}$ based on the patients $\{i : R_i = 1\}$ with observed stratum, and estimate $h(W_i, \mathbf{b})$ by $h(W_i, \widehat{\mathbf{b}})$. An alternative solution is, for example, to use the less restrictive local logistic regression approach (e.g., Loader (1999)).

Similarly, various approaches may be used to estimate $\mu(G_i, t) = 1/\mathbb{P}(C \geq t|G_i)$. These include parametric and non-parametric methods.

The choice between competing estimation methods may be guided by the investigator's prior knowledge of the relationship between S and W, and of the censoring mechanism. In our simulation study, we considered the least favorable case where both $\mathbb{P}(S_i = l|W_i)$ and $\mu(G_i, \cdot)$ are unknown. We estimated the $\mathbb{P}(S_i = l|W_i)$ by local logistic regression, using the `R` package `locfit` (available at `http://cran.r-project.org/web/packages/locfit/`). We used non-parametric Kaplan-Meier estimators within each treatment group to estimate the censoring survival functions $\mathbb{P}(C \geq t|G_i)$.

For each configuration of the design parameters, 1000 replications were obtained using the software `R`. Based on these 1000 repetitions, we obtained the empirical size (case (a)) and power (cases (b) and (c)) of the "estimated version" $\widehat{U}$ of the proposed test $\widetilde{U}$, at the significance level 0.05. For comparison, we included the results of the stratified log-rank test based on complete cases only (*i.e.* on individuals with known stratum). In the sequel, we shall refer this latter test to as U_{cc} for short. Table 2 summarizes the results for an overall stratum missingness percentage equal to 40% (the results for 20% are similar and are therefore not presented).

Table 2. Empirical size and power of $\widehat{U}$ and U_{cc}, based on 1000 replicates.

		Censoring percentages (c_1, c_2)							
		(5,20)		(20,50)		(30,20)		(40,10)	
n	(r_1, r_2)	$\widehat{U}$	U_{cc}	$\widehat{U}$	U_{cc}	$\widehat{U}$	U_{cc}	$\widehat{U}$	U_{cc}
100	(1, 1)	.067	.045	.070	.067	.057	.045	.061	.057
	(1.5, 1.5)	.354	.228	.272	.204	.401	.258	.360	.221
	(1.25, 2)	.349	.227	.303	.220	.361	.231	.314	.216
200	(1, 1)	.074	.055	.065	.051	.067	.049	.051	.047
	(1.5, 1.5)	.692	.440	.565	.369	.651	.381	.574	.423
	(1.25, 2)	.592	.386	.605	.418	.603	.403	.465	.401
300	(1, 1)	.060	.046	.076	.048	.078	.052	.073	.048
	(1.5, 1.5)	.823	.596	.762	.472	.822	.566	.698	.568
	(1.25, 2)	.772	.535	.813	.527	.746	.522	.611	.583

From these results, it appears that the proposed test $\widehat{U}$ performs well and clearly outperforms the stratified test based on the complete cases (which was the only alternative test available so far for the problem considered in this paper). The empirical level of $\widehat{U}$ tends to exceed (but only slightly) 0.05 in all cases. This may be due to the replacement of the unknown $\mathbb{P}(S_i = l|W_i)$ and $\mu(G_i, \cdot)$ by their estimations, which causes the null asymptotic distribution of $\widehat{U}$ to be slightly distorted from the $\chi^2(K-1)$. But as expected, in cases (b) and (c), the powers of $\widehat{U}$ are greater than those of U_{cc} for every sample size and censoring percentages. In particular, $\widehat{U}$ maintains a high power even when the censoring percentage heavily depends on the treatment group (20% in group 1, 50% in group 2), while at the same time the powers of U_{cc} substantially decrease.

4. Discussion

We have constructed and investigated a modified version of the stratified log-rank test of no randomized treatment effect. This new test statistic is useful when the stratum information is missing at random for some patients and the distribution of the censoring time depends on the treatment group. From our simulations, we have found that this test performs well compared to the only alternative available so far, namely a complete-case based stratified log-rank test. Now, several questions still deserve attention.

First, we have assumed a missing-at-random mechanism for the stratum missingness. Investigating the robustness of the proposed test to a deviation to this assumption constitutes a topic for further numerical investigations. Extending the proposed method to non-ignorable missingness may be a non-trivial task however: the missing-at-random assumption is central in our proofs and for estimating the stratum belonging probabilities $\mathbb{P}(S_i = l|W_i)$ when they are unknown. One may also investigate the case where the stratum missingness depends on the treatment group.

Second, in order to accommodate group-dependent censoring, we have used the inverse probability of censoring weighted principle, with weight function $\mu(G_i, t) = h(\mathbb{P}(C \geq t|G_i))$ and $h(x) = 1/x$. Alternative test statistics $\widetilde{Z}_k$ may be obtained by choosing other forms for h (we refer to DiRienzo and Lagakos (2001) for alternative forms of weighting functions in the different context of bias correction for score tests arising from misspecified proportional hazards regression models). Searching for the function h which yields the most efficient testing procedure constitutes another non-trivial but very interesting task.

Another interesting future research direction is as follows. The stratified log-rank test can be viewed as a score test in a stratified Cox regression model. Therefore, the problem we considered in this paper can be viewed as the problem of implementing a score test of no randomized treatment effect in a stratified Cox model with missing stratum information. In the past two decades, a large amount of literature has been devoted to the problems of estimation and testing in the unstratified proportional hazards Cox model with missing covariates, but this literature has essentially focused on the case where the missing covariate has a proportional effect on the hazard of failure. As far as we know, the case where the missing covariate has a non-proportional effect and therefore is used to stratify was only few investigated. Dupuy and Leconte (2008, 2009) have considered this problem, but no treatment-dependent censoring was assumed. We hope that the present work can constitute a first step in the direction of relaxing this hypothesis.

Appendix A. Proof of Lemma 1.

We have

$$\begin{aligned} n^{-\frac{1}{2}}\widetilde{Z}_k &= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^\tau \mu(G_i,t)G_i^k\,dN_i(t) - n^{-\frac{1}{2}}\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau \mu(G_i,t)D_i^l\widetilde{E}_{k,l}^{(n)}(t)\,dN_i(t) \\ &:= A_{1,k}^{(n)} - A_{2,k}^{(n)}, \end{aligned}$$

and $A_{2,k}^{(n)}$ can be written as

$$\begin{aligned} A_{2,k}^{(n)} = n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau \widetilde{e}_{k,l}(t)\left[\frac{1}{n}\sum_{i=1}^n \mu(G_i,t)D_i^l dN_i(t) - \kappa_l(t)dt\right] \\ +n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau v_{k,l}^{(n)}(t)\left[\frac{1}{n}\sum_{i=1}^n \mu(G_i,t)D_i^l dN_i(t) - \kappa_l(t)dt\right] + n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau \widetilde{E}_{k,l}^{(n)}(t)\kappa_l(t)dt, \end{aligned}$$

where $v_{k,l}^{(n)}(t) := \widetilde{E}_{k,l}^{(n)}(t) - \widetilde{e}_{k,l}(t)$. We let the first term in $A_{2,k}^{(n)}$ unchanged. The second and third terms satisfy respectively the following two technical lemmas, whose proofs are postponed to the Appendix B:

Lemma 2. *Under conditions C1-C4,*

$$n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau v_{k,l}^{(n)}(t)\left[\frac{1}{n}\sum_{i=1}^n \mu(G_i,t)D_i^l dN_i(t) - \kappa_l(t)dt\right] \xrightarrow{p} 0 \text{ as } n\to\infty.$$

Lemma 3. *Under conditions C1-C4, as $n\to\infty$,*

$$n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau \widetilde{E}_{k,l}^{(n)}(t)\kappa_l(t)dt = n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau \left\{\widetilde{e}_{k,l}(t) + \frac{\widetilde{S}_{k,l}^{(n)}(t)}{\widetilde{s}_l(t)} - \frac{\widetilde{s}_{k,l}(t)\widetilde{S}_l^{(n)}(t)}{\widetilde{s}_l(t)^2}\right\}\kappa_l(t)dt + o_p(1).$$

Using these lemmas, we can re-write $n^{-\frac{1}{2}}\widetilde{Z}_k$ as: $n^{-\frac{1}{2}}\widetilde{Z}_k =$

$$\begin{aligned} n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^\tau \mu(G_i,t)G_i^k\,dN_i(t) - n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau \widetilde{e}_{k,l}(t)\left[\frac{1}{n}\sum_{i=1}^n \mu(G_i,t)D_i^l dN_i(t) - \kappa_l(t)dt\right] \\ -n^{\frac{1}{2}}\sum_{l=1}^L\int_0^\tau \frac{1}{\widetilde{s}_l(t)}\left\{\widetilde{s}_{k,l}(t) + \widetilde{S}_{k,l}^{(n)}(t) - \frac{\widetilde{s}_{k,l}(t)\widetilde{S}_l^{(n)}(t)}{\widetilde{s}_l(t)}\right\}\kappa_l(t)dt + o_p(1). \end{aligned}$$

Rearranging the terms in this expression concludes the proof of the first statement in Lemma 1. We now turn to the expectation of $Q_{i,k}$ under H_0.

For every $i = 1,\ldots,n$, $k = 1,\ldots,K$, and $t\in[0,\tau]$, $\mu(G_i,t)G_i^k$ is $\mathcal{H}_{t,i}$-measurable, hence the process $(\int_0^t \mu(G_i,s)G_i^k\,dM_i(s))_{t\geq 0}$ is a zero-mean martingale. It follows that

$$\begin{aligned} \mathbb{E}\left[\int_0^\tau\sum_{l=1}^L D_i^l\mu(G_i,t)G_i^k\,dN_i(t)\right] &= \mathbb{E}\left[\int_0^\tau \mu(G_i,t)G_i^k\,dN_i(t)\right] \\ &= \int_0^\tau \mathbb{E}\left[\mu(G_i,t)G_i^kY_i(t)\zeta_i(t)\right]\,dt, \end{aligned}$$

where the first equality follows by noting that $\sum_{l=1}^{L} D_i^l = 1$. Similarly, we have

$$\mathbb{E}\left[\int_0^\tau \sum_{l=1}^{L} D_i^l \mu(G_i,t)\widetilde{e}_{k,l}(t)\, dN_i(t)\right] = \sum_{l=1}^{L}\int_0^\tau \widetilde{e}_{k,l}(t)\kappa_l(t)\, dt,$$

$$\mathbb{E}\left[\int_0^\tau \sum_{l=1}^{L} D_i^l \mu(G_i,t) G_i^k Y_i(t)\frac{\kappa_l(t)}{\widetilde{s}_l(t)}\, dt\right] = \sum_{l=1}^{L}\int_0^\tau \widetilde{e}_{k,l}(t)\kappa_l(t)\, dt,$$

and

$$\mathbb{E}\left[\int_0^\tau \sum_{l=1}^{L} D_i^l \mu(G_i,t)\widetilde{e}_{k,l}(t) Y_i(t)\frac{\kappa_l(t)}{\widetilde{s}_l(t)}\, dt\right] = \sum_{l=1}^{L}\int_0^\tau \widetilde{e}_{k,l}(t)\kappa_l(t)\, dt.$$

Thus

$$\begin{aligned}\mathbb{E}[Q_{i,k}] &= \int_0^\tau \mathbb{E}\left[\mu(G_i,t)G_i^k Y_i(t)\zeta_i(t)\right]\, dt - \sum_{l=1}^{L}\int_0^\tau \widetilde{e}_{k,l}(t)\kappa_l(t)\, dt \\ &:= B_{1,k} - B_{2,k}.\end{aligned}$$

We now prove that under H_0, $B_{2,k} = \int_0^\tau \mathbb{E}[G^k]\mathbb{E}[Y(t)\zeta(t)\mu(G,t)]\, dt$. First, remark that

$$\begin{aligned}\widetilde{s}_{k,l}(t) &= \mathbb{E}\left[Y(t)G^k D^l \mu(G,t)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[Y(t)G^k D^l \mu(G,t)|G,S,W,R\right]\right] \\ &= \mathbb{E}\left[G^k D^l \mu(G,t)\mathbb{E}\left[Y(t)|G,S,W,R\right]\right] \\ &\overset{H_0}{=} \mathbb{E}\left[G^k D^l \mu(G,t)\mathbb{E}\left[1(T^0 \geq t)|S\right]\mathbb{E}\left[1(C \geq t)|G\right]\right]\end{aligned}$$

where the third to last line follows from the assumptions C1 and from the fact that under H_0, the distribution of T^0 does not depend on G. Then, by the independence of G and (S,W,R), we get that

$$\begin{aligned}\widetilde{s}_{k,l}(t) &= \mathbb{E}\left[G^k D^l \mathbb{E}\left[1(T^0 \geq t)|S\right]\right] \\ &= \mathbb{E}\left[G^k\right]\mathbb{E}\left[D^l \mathbb{E}\left[1(T^0 \geq t)|S\right]\right].\end{aligned}$$

Finally, using the same arguments as above, and the properties of conditional expectation, we have:

$$\begin{aligned}\widetilde{s}_{k,l}(t) &= \mathbb{E}\left[G^k\right]\mathbb{E}\left[D^l \mathbb{E}\left[1(T^0 \geq t)|S\right]\mu(G,t)\mathbb{E}\left[1(C \geq t)|G\right]\right] \\ &= \mathbb{E}\left[G^k\right]\mathbb{E}\left[D^l \mu(G,t)\mathbb{E}\left[Y(t)|G,S,W,R\right]\right] \\ &= \mathbb{E}\left[G^k\right]\widetilde{s}_l(t).\end{aligned}$$

Thus, under H_0,

$$\begin{aligned}B_{2,k} &= \sum_{l=1}^{L}\int_0^\tau \mathbb{E}\left[G^k\right]\kappa_l(t)\, dt \\ &= \int_0^\tau \mathbb{E}\left[G^k\right]\mathbb{E}[Y(t)\zeta(t)\mu(G,t)]\, dt.\end{aligned}$$

Using similar arguments, we get that $B_{1,k} = \int_0^\tau \mathbb{E}\left[G^k\right]\mathbb{E}[Y(t)\zeta(t)\mu(G,t)]\, dt$ under H_0 and thus, $\mathbb{E}[Q_{i,k}] = 0$.

□

Appendix B. Proofs of Lemma 2 and Lemma 3.

Proof of Lemma 2. Let $l \in \{1, \dots, L\}$, and decompose

$$n^{\frac{1}{2}} \int_0^\tau v_{k,l}^{(n)}(t) \left[\frac{1}{n} \sum_{i=1}^n \mu(G_i, t) D_i^l dN_i(t) - \kappa_l(t) dt \right]$$

as $C_{1,k,l}^{(n)} + C_{2,k,l}^{(n)}$, where

$$C_{1,k,l}^{(n)} := n^{\frac{1}{2}} \int_0^\tau v_{k,l}^{(n)}(t) \left[\frac{1}{n} \sum_{i=1}^n \mu(G_i, t) D_i^l dN_i(t) - \frac{1}{n} \sum_{i=1}^n \mu(G_i, t) D_i^l Y_i(t) \zeta_i(t)\, dt \right]$$

and

$$C_{2,k,l}^{(n)} := \int_0^\tau v_{k,l}^{(n)}(t) \cdot n^{\frac{1}{2}} \left[\frac{1}{n} \sum_{i=1}^n \mu(G_i, t) D_i^l Y_i(t) \zeta_i(t)\, dt - \kappa_l(t) dt \right].$$

In the following development, we first show that $C_{1,k,l}^{(n)} \xrightarrow{p} 0$ as $n \to \infty$. Note that $C_{1,k,l}^{(n)}$ is of the form:

$$C_{1,k,l}^{(n)} = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau H_{i,k,l}^{(n)}(t) dM_i(t),$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(s)\zeta_i(s) ds$ and $H_{i,k,l}^{(n)}(t) := v_{k,l}^{(n)}(t) \mu(G_i, t) D_i^l$ is a predictable process with respect to $\mathcal{H}_t := \bigvee_{i=1}^n \mathcal{H}_{t,i}$. Moreover, $H_{i,k,l}^{(n)}(t)$ is bounded on $[0, \tau]$ since $|H_{i,k,l}^{(n)}(t)| \le |v_{k,l}^{(n)}(t)| \cdot \frac{1}{c_0} \le \frac{2}{c_0}$. Define the process $(C_{1,k,l}^{(n)}(t))_{t \ge 0}$ by

$$C_{1,k,l}^{(n)}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t H_{i,k,l}^{(n)}(s) dM_i(s).$$

Then $(C_{1,k,l}^{(n)}(t))_{t \ge 0}$ is an $\mathcal{H}_t$-martingale, and $C_{1,k,l}^{(n)} := C_{1,k,l}^{(n)}(\tau)$. Now, the predictable variation process $< C_{1,k,l}^{(n)} > (t)$ of $C_{1,k,l}^{(n)}(t)$ is

$$< C_{1,k,l}^{(n)} > (t) = \int_0^t \frac{1}{n} \sum_{i=1}^n \left\{ H_{i,k,l}^{(n)}(s) \right\}^2 Y_i(s) \zeta_i(s) ds := \int_0^t X^{(n)}(s) ds.$$

Under the regularity conditions stated in Section 2, it is not difficult to check that $X^{(n)}(s) \xrightarrow{p} 0$ as $n \to \infty$ for every $s \in [0, \tau]$. Moreover, for any $s \in [0, \tau]$ and $n \ge 1$,

$$\begin{aligned} |X^{(n)}(s)| &= \left| \frac{1}{n} \sum_{i=1}^n \left\{ H_{i,k,l}^{(n)}(s) \right\}^2 Y_i(s) \zeta_i(s) \right| \\ &\le \frac{1}{n} \sum_{i=1}^n \left\{ H_{i,k,l}^{(n)}(s) \right\}^2 \zeta_i(s) \\ &\le \frac{4}{c_0^2} c_1 \end{aligned}$$

hence by Proposition II.5.3 of Andersen *et al.* (1993), $< C_{1,k,l}^{(n)} > (t) \xrightarrow{p} 0$ as $n \to \infty$. Next, for any $\epsilon > 0$ and $t \in [0, \tau]$, let

$$C_{1,k,l,\epsilon}^{(n)}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t H_{i,k,l}^{(n)}(s) 1(n^{-\frac{1}{2}} |H_{i,k,l}^{(n)}(s)| \geq \epsilon) dM_i(s).$$

Then the nonnegative $< C_{1,k,l,\epsilon}^{(n)} > (t)$ satisfies the following:

$$\begin{aligned} < C_{1,k,l,\epsilon}^{(n)} > (t) &= \int_0^t \frac{1}{n} \sum_{i=1}^{n} \left\{ H_{i,k,l}^{(n)}(s) \right\}^2 1(n^{-\frac{1}{2}} |H_{i,k,l}^{(n)}(s)| \geq \epsilon) Y_i(s) \zeta_i(s) ds \\ &\leq \ < C_{1,k,l}^{(n)} > (t), \end{aligned}$$

and therefore $< C_{1,k,l,\epsilon}^{(n)} > (t) \xrightarrow{p} 0$ as $n \to \infty$. It follows (see Theorem 5.1.1 in Fleming and Harrington (1991)) that $C_{1,k,l}^{(n)}(t) \xrightarrow{d} 0$ as $n \to \infty$ for every $t \in [0, \tau]$. Therefore, $C_{1,k,l}^{(n)}(t) \xrightarrow{p} 0$ as $n \to \infty$ and in particular, if $t = \tau$, $C_{1,k,l}^{(n)} \xrightarrow{p} 0$ as $n \to \infty$.

We now prove that $C_{2,k,l}^{(n)} \xrightarrow{p} 0$ as $n \to \infty$. To this end, we show that the integrand of $C_{2,k,l}^{(n)}$ converges to 0 uniformly in probability on $[0, \tau]$. We have:

$$\begin{aligned} &\sup_{t \in [0,\tau]} \left| v_{k,l}^{(n)}(t) \cdot n^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} \mu(G_i, t) D_i^l Y_i(t) \zeta_i(t) - \kappa_l(t) \right) \right| \\ &\qquad \leq \sup_{t \in [0,\tau]} \left| v_{k,l}^{(n)}(t) \right| \cdot \sup_{t \in [0,\tau]} \left| n^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} \mu(G_i, t) D_i^l Y_i(t) \zeta_i(t) - \kappa_l(t) \right) \right|. \end{aligned} \tag{4}$$

Somewhat straightforward Glivenko-Cantelli arguments yield that $\widetilde{E}_{k,l}^{(n)}(t)$ is uniformly consistent (on $[0, \tau]$) for $\widetilde{e}_{k,l}(t)$. It follows that as $n \to \infty$,

$$\sup_{t \in [0,\tau]} \left| v_{k,l}^{(n)}(t) \right| \xrightarrow{p} 0. \tag{5}$$

Then, the classes $\{Y(t), t \in [0, \tau]\}$ and $\{\mu(G, t), t \in [0, \tau]\}$ are both Donsker (by Lemma 4.1 in Kosorok (2008)), and thus so is $\{\mu(G, t) Y(t), t \in [0, \tau]\}$ since products of bounded Donsker classes are Donsker. Similarly, $\{\zeta(t), t \in [0, \tau]\}$ is Donsker (Corollary 9.32 in Kosorok (2008)) and $\{D^l\}$ is Donsker. Finally, the class $\{\mu(G, t) D^l Y(t) \zeta(t), t \in [0, \tau]\}$ is Donsker (as a product of bounded Donsker classes) and therefore, the process $G_n(\cdot) := n^{\frac{1}{2}} \left(n^{-1} \sum_{i=1}^{n} \mu(G_i, \cdot) D_i^l Y_i(\cdot) \zeta_i(\cdot) - \kappa_l(\cdot) \right)$ converges weakly to some mean zero Gaussian process G. By the continuous mapping theorem, $\sup_{t \in [0,\tau]} |G_n(t)|$ converges weakly to $\sup_{t \in [0,\tau]} |G(t)|$ and therefore $\sup_{t \in [0,\tau]} |G_n(t)| = O_p(1)$. Combining this result and result (5), in (4), yields that

$$\sup_{t \in [0,\tau]} \left| v_{k,l}^{(n)}(t) \cdot n^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} \mu(G_i, t) D_i^l Y_i(t) \zeta_i(t) - \kappa_l(t) \right) \right| \xrightarrow{p} 0$$

as $n \to \infty$. This implies that $C_{2,k,l}^{(n)} \xrightarrow{p} 0$ as $n \to \infty$.

Hence, for any $l \in \{1, \ldots, L\}$, $C_{1,k,l}^{(n)} + C_{2,k,l}^{(n)} \xrightarrow{p} 0$ as $n \to \infty$, and thus $\sum_{l=1}^{L} C_{1,k,l}^{(n)} + C_{2,k,l}^{(n)} \xrightarrow{p} 0$. This concludes the proof of Lemma 2.

□

Proof of Lemma 3. Decompose

$$n^{\frac{1}{2}} \sum_{l=1}^{L} \int_0^{\tau} \widetilde{E}_{k,l}^{(n)}(t) \kappa_l(t) dt$$

as

$$n^{\frac{1}{2}} \sum_{l=1}^{L} \int_0^{\tau} \left\{ \widetilde{e}_{k,l}(t) + \frac{\widetilde{S}_{k,l}^{(n)}(t)}{\widetilde{s}_l(t)} - \frac{\widetilde{s}_{k,l}(t) \widetilde{S}_l^{(n)}(t)}{\widetilde{s}_l(t)^2} \right\} \kappa_l(t) dt$$
$$+ \sum_{l=1}^{L} \int_0^{\tau} \left\{ \frac{\widetilde{S}_{k,l}^{(n)}(t)}{\widetilde{s}_l(t)} - \frac{\widetilde{s}_{k,l}(t) \widetilde{S}_l^{(n)}(t)}{\widetilde{s}_l(t)^2} \right\} \cdot n^{\frac{1}{2}} \left(\frac{\widetilde{s}_l(t)}{\widetilde{S}_l^{(n)}(t)} - 1 \right) \kappa_l(t) dt. \quad (6)$$

The second term on the right-hand side of (6) can be shown to converge to 0 in probability as $n \to \infty$. The arguments are similar to the ones used in the proof of Lemma 2 and are therefore omitted.

□

References

Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N., 1993. Statistical models based on counting processes. Springer: New York.

Cain, L.E., Cole, S.R., 2009. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. *Statistics in Medicine* 28, 1725-1738.

Carroll, R.J., Ruppert, D., Stefanski, L.A., 1995. Measurement Error in Nonlinear Models. Monographs on Statistics and Applied Probability vol. 63, Chapman & Hall: London.

DiRienzo, A.G., Lagakos, S.W., 2001. Bias Correction for Score Tests Arising from Misspecified Proportional Hazards Regression Models. *Biometrika* 88, 421–434.

Dupuy, J.-F., Leconte, E., 2008. Cox regression with missing values of a covariate having a non-proportional effect on hazard of failure. In: Mathematical methods in survival analysis, reliability and quality of life (Applied Stochastic Methods Series), 133-150, ISTE: London.

Dupuy, J.-F., Leconte, E., 2009. A study of regression calibration in a partially observed stratified Cox model. *Journal of Statistical Planning and Inference* 139, 317–328.

Fleming, T.R., Harrington, D.P., 1991. Counting processes and survival analysis. John Wiley & Sons: New York.

Klein, J.P., Moeschberger, M.L., 1997. Survival Analysis: Techniques for Censored and Truncated Data. Springer: New York.

Kosorok, M.R., 2008. Introduction to empirical processes and semiparametric inference. Springer: New York.

Loader, C., 1999. Local Regression and Likelihood. Springer: New York.

Martinussen, T., Scheike, T.H., 2006. Dynamic Regression Models for Survival Data. Springer: New York.

Robins, J.M., Finkelstein, D.M., 2000. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 56, 779–788.

Thurston, S. W., Spiegelman, D., Ruppert, D., 2003. Equivalence of regression calibration methods in main study/external validation study designs. *Journal of Statistical Planning and Inference* 113, 527–539.

Weller, E.A., Milton, D.K., Eisen, E.A., Spiegelman, D., 2007. Regression calibration for logistic regression with multiple surrogates for one exposure. *Journal of Statistical Planning and Inference* 137, 449–461.

Yoshida, M., Matsuyama, Y., Ohashi, Y., 2007. Estimation of treatment effect adjusting for dependent censoring using the IPCW method: an application to a large primary prevention study for coronary events (MEGA study). *Clinical Trials* 4, 318–328.